

IN THE CLAIMS

1. (Currently Amended) A method for determining dominant phrase vectors in a topological vector space for a semantic content of a document on a computer system, the method comprising:

identifying a directed set of concepts as a dictionary, the directed set including a maximal element and at least one concept, and at least one chain from the maximal element to every concept;

selecting a subset of the chains to form a basis for the dictionary;

accessing dominant phrases for the document, the dominant phrases representing a condensed content for the document;

measuring how concretely each dominant phrase is represented in each chain in the basis and the dictionary;

constructing at least one state vector in the topological vector space for each dominant phrase using a-the measures of how concretely each dominant phrase is represented in each chain in the dictionary and a-the basis; and

collecting the state vectors into the dominant phrase vectors for the document.

2. (Original) A method according to claim 1, wherein accessing dominant phrases includes extracting the dominant phrases from the document using a phrase extractor.

3. (Original) A method according to claim 1, wherein accessing dominant phrases includes storing the dominant phrases in computer memory accessible by the computer system.

4. (Original) A method according to claim 1, the method further comprising forming a semantic abstract comprising the dominant phrase vectors.

5. (Currently Amended) A method for determining dominant vectors in a topological vector space for a semantic content of a document on a computer system, the method comprising:

identifying a directed set of concepts as a dictionary, the directed set including a maximal element and at least one concept, and at least one chain from the maximal element to every concept;

selecting a subset of the chains to form a basis for the dictionary;

Serial No.: 09/615,726

2

Docket No.: 6647-13

storing the document in computer memory accessible by the computer system;
extracting words from at least a portion of the document;
measuring how concretely each word is represented in each chain in the basis and the
dictionary;

constructing a state vector in the topological vector space for each word using ~~a~~the
measures of how concretely each word is represented in each chain in the dictionary and a~~the~~
basis;

filtering the state vectors; and

collecting the filtered state vectors into the dominant vectors for the document.

6. (Original) A method according to claim 5, wherein extracting words includes extracting words from the entire document.

7. (Original) A method according to claim 5, wherein filtering the state vectors includes selecting the state vectors that occur with highest frequencies.

8. (Original) A method according to claim 5, wherein filtering the state vectors includes:

calculating a centroid in the topological vector space for the state vectors; and
selecting the state vectors nearest the centroid.

9. (Original) A method according to claim 5, the method further comprising forming a semantic abstract comprising the dominant vectors.

10. (Original) A computer-readable medium containing a program to determine dominant vectors in a topological vector space for a semantic content of a document on a computer system, the program being executable on the computer system to implement the method of claim 5.

11. (Currently Amended) A method for determining a semantic abstract in a topological vector space for a semantic content of a document on a computer system, the method comprising:

identifying a directed set of concepts as a dictionary, the directed set including a maximal element and at least one concept, and at least one chain from the maximal element to every concept;

selecting a subset of the chains to form a basis for the dictionary;

storing the document in computer memory accessible by the computer system;

determining dominant ~~phrase vectors~~ phrases for the document;

measuring how concretely each dominant phrase is represented in each chain in the basis and the dictionary;

constructing dominant phrase vectors in the topological vector space for the dominant phrases using the measures of how concretely each dominant phrase is represented in each chain in the dictionary and the basis;

~~determining dominant vectors~~ selecting words for the document; and

measuring how concretely each word is represented in each chain in the basis and the dictionary;

constructing dominant vectors in the topological vector space for the words using the measures of how concretely each word is represented in each chain in the dictionary and the basis; and

generating the semantic abstract using the dominant phrase vectors and the dominant vectors.

12. (Original) A method according to claim 11, wherein generating the semantic abstract includes reducing the dominant phrase vectors based on the dominant vectors.

13. (Original) A method according to claim 11, wherein generating the semantic abstract includes reducing the dominant vectors based on the dominant phrase vectors.

14. (Original) A method according to claim 11, wherein generating the semantic abstract includes obtaining a probability distribution function for a reduced set of the dominant phrase vectors similar to a probability distribution function for the dominant phrase vectors.

15. (Original) A method according to claim 11, the method further comprising identifying the lexemes or lexeme phrases corresponding to state vectors in the semantic abstract.

16. (Original) A computer-readable medium containing a program to determine a semantic abstract in a topological vector space for a semantic content of a document on a computer system, the program being executable on the computer system to implement the method of claim 11.

17. (Currently Amended) A method for comparing the semantic content of first and second documents on a computer system, the method comprising:
identifying a directed set of concepts as a dictionary, the directed set including a maximal element and at least one concept, and at least one chain from the maximal element to every concept;
selecting a subset of the chains to form a basis for the dictionary;
accessing dominant phrases for the first document, the dominant phrases representing a condensed content for the first document;
measuring how concretely each dominant phrase for the first document is represented in each chain in the basis and the dictionary;
constructing at least one state vector for the first document in the topological vector space for each dominant phrase for the first document using the measures of how concretely each dominant phrase for the first document is represented in each chain in the dictionary and the basis;
collecting the state vectors for the first document into the semantic abstract for the first document;
determining a semantic abstracts abstract for the first and second documents;
document;
measuring a distance between the semantic abstracts; and
classifying how closely related the first and second documents are using the distance.

18. (Original) A method according to claim 17, wherein measuring a distance includes measuring a Hausdorff distance between the semantic abstracts.

19. (Original) A method according to claim 17, wherein measuring a distance includes determining a centroid vector in the topological vector space for each semantic abstract.

20. (Original) A method according to claim 19, wherein measuring a distance further includes measuring an angle between the centroid vectors.

21. (Original) A method according to claim 19, wherein measuring a distance further includes measuring a Euclidean distance between the centroid vectors.

22. (Original) A computer-readable medium containing a program to compare the semantic content of first and second documents on a computer system, the program being executable on the computer system to implement the method of claim 17.

23. (Currently Amended) A method for locating a second document on a computer with a semantic content similar to a first document, the method comprising:

- ~~determining a semantic abstract for the first document;~~
- identifying a directed set of concepts as a dictionary, the directed set including a maximal element and at least one concept, and at least one chain from the maximal element to every concept;
- selecting a subset of the chains to form a basis for the dictionary;
- accessing dominant phrases for the first document, the dominant phrases representing a condensed content for the first document;
- measuring how concretely each dominant phrase for the first document is represented in each chain in the basis and the dictionary;
- constructing at least one state vector for the first document in the topological vector space for each dominant phrase for the first document using the measures of how concretely each dominant phrase for the first document is represented in each chain in the dictionary and the basis;
- collecting the state vectors for the first document into the semantic abstract for the first document;
- locating a second document;
- determining a semantic abstract for the second document;

measuring a distance between the semantic abstracts for the first and second documents;
classifying how closely related the first and second documents are using the distance;
and
if the second document is classified as having a semantic content similar to the semantic content of the first document, selecting the second document.

24. (Original) A method according to claim 23, the method further comprising, if the second document is classified as not having a semantic content similar to the semantic content of the first document, rejecting the second document.

25. (Currently Amended) An apparatus on a computer system to determine a semantic abstract in a topological vector space for a semantic content of a document stored on the computer system, the apparatus comprising:
a phrase extractor adapted to extract phrases from the document;
a state vector constructor adapted to construct ~~at least one~~ state vectors in the topological vector space for each phrase extracted by the phrase extractor; extractor. the state vectors measuring how concretely each phrase extracted by the phrase extractor is represented in each chain in a basis and a dictionary, the dictionary including a directed set of concepts including a maximal element and at least one chain from the maximal element to every concept in the directed set, the basis including a subset of chains in the directed set; and
collection means for collecting the state vectors into the semantic abstract for the document.

26. (Original) An apparatus according to claim 25, the apparatus further comprising filter means for filtering the state vectors to reduce the size of the semantic abstract.

27. (Original) An apparatus according to claim 25, wherein the state vector constructor is further adapted to construct a state vector for each word in the document.

28. (Canceled)

29. (Currently Amended) A method for determining a semantic abstract in a topological vector space for a semantic content of a document on a computer system, the method comprising:

extracting dominant phrases from the document using a phrase extractor, the dominant phrases representing a condensed content for the document;

identifying a directed set of concepts as a dictionary, the directed set including a maximal element and at least one concept, and at least one chain from the maximal element to every concept;

selecting a subset of the chains to form a basis for the dictionary;

measuring how concretely each dominant phrase is represented in each chain in the basis and the dictionary;

constructing at least one first state vector in the topological vector space for each dominant phrase using a-the measures of how concretely each dominant phrase is represented in each chain in the dictionary and a-the basis;

collecting the first state vectors into dominant phrase vectors for the document;

extracting words from at least a portion of the document;

constructing a second state vector in the topological vector space for each word using the dictionary and the basis;

filtering the second state vectors;

collecting the filtered second state vectors into dominant vectors for the document;

and

generating the semantic abstract using the dominant phrase vectors and the dominant vectors.

30. (Original) A method according to claim 29, the method further comprising comparing the semantic abstract with a second semantic abstract for a second document to determine how closely related the contents of the documents are.

31. (New) A method according to claim 1, wherein:

measuring how concretely each dominant phrase is represented in each chain in the basis and the dictionary includes:

identifying at least one lexeme in each dominant phrase; and

measuring how concretely each lexeme in each dominant phrase is represented in each chain in the basis and the dictionary; and

constructing at least one state vector in the topological vector space for each dominant phrase includes constructing at least one state vector in the topological vector space for each lexeme in each dominant phrase using the measures of how concretely each lexeme in each dominant phrase is represented in each chain in the basis and the dictionary.

32. (New) A method according to claim 17, wherein determining a semantic abstract for the second document includes:

accessing dominant phrases for the second document, the dominant phrases representing a condensed content for the second document;

measuring how concretely each dominant phrase for the second document is represented in each chain in the basis and the dictionary;

constructing at least one state vector for the second document in the topological vector space for each dominant phrase for the second document using the measures of how concretely each dominant phrase for the second document is represented in each chain in the dictionary and the basis; and

collecting the state vectors for the second document into the semantic abstract for the second document.

33. (New) A method according to claim 23, wherein determining a semantic abstract for the second document includes:

accessing dominant phrases for the second document, the dominant phrases representing a condensed content for the second document;

measuring how concretely each dominant phrase for the second document is represented in each chain in the basis and the dictionary;

constructing at least one state vector for the second document in the topological vector space for each dominant phrase for the second document using the measures of how concretely each dominant phrase for the second document is represented in each chain in the dictionary and the basis; and

collecting the state vectors for the second document into the semantic abstract for the second document.

34. (New) A method according to claim 29, wherein constructing a second state vector in the topological vector space for each word using the dictionary and the basis includes:

Serial No.: 09/615,726

9

Docket No.: 6647-13

measuring how concretely each word is represented in each chain in the basis and the dictionary; and

constructing the second state vectors in the topological vector space for each word using the measures of how concretely each word is represented in each chain in the dictionary and the basis.

35. (New) A method according to claim 1, wherein constructing at least one state vector includes constructing the state vectors in the topological vector space for each dominant phrase using the measures of how concretely each dominant phrase is represented in each chain in the dictionary and the basis, the state vectors independent of the document.

36. (New) A method according to claim 5, wherein constructing a state vector includes constructing the state vector in the topological vector space for each word using the measures of how concretely each word is represented in each chain in the dictionary and the basis, the state vectors independent of the document.

37. (New) A method according to claim 11, wherein constructing dominant vectors includes constructing dominant vectors in the topological vector space for the words using the measures of how concretely each word is represented in each chain in the dictionary and the basis, the dominant vectors independent of the document.

38. (New) A method according to claim 17, wherein constructing at least one state vector includes constructing the state vectors for the first document in the topological vector space for each dominant phrase for the first document using the measures of how concretely each dominant phrase for the first document is represented in each chain in the dictionary and the basis, the state vectors independent of the first document and the second document.

39. (New) A method according to claim 12, wherein constructing at least one state vector includes constructing at least one state vector for the first document in the topological vector space for each dominant phrase for the first document using the measures of how concretely each dominant phrase for the first document is represented in each chain in the dictionary and the basis, the state vectors independent of the first document and the second document.

40. (New) An apparatus according to claim 25, wherein the state vector constructor is operative to construct the state vectors independent of the document.

41. (New) A method according to claim 29, wherein:

constructing at least one first state vector includes constructing the first state vector in the topological vector space for each dominant phrase using the measures of how concretely each dominant phrase is represented in each chain in the dictionary and the basis, the first state vector independent of the document; and

constructing a second state vector includes constructing the second state vector in the topological vector space for each word using the dictionary and the basis, the second state vector independent of the document.